



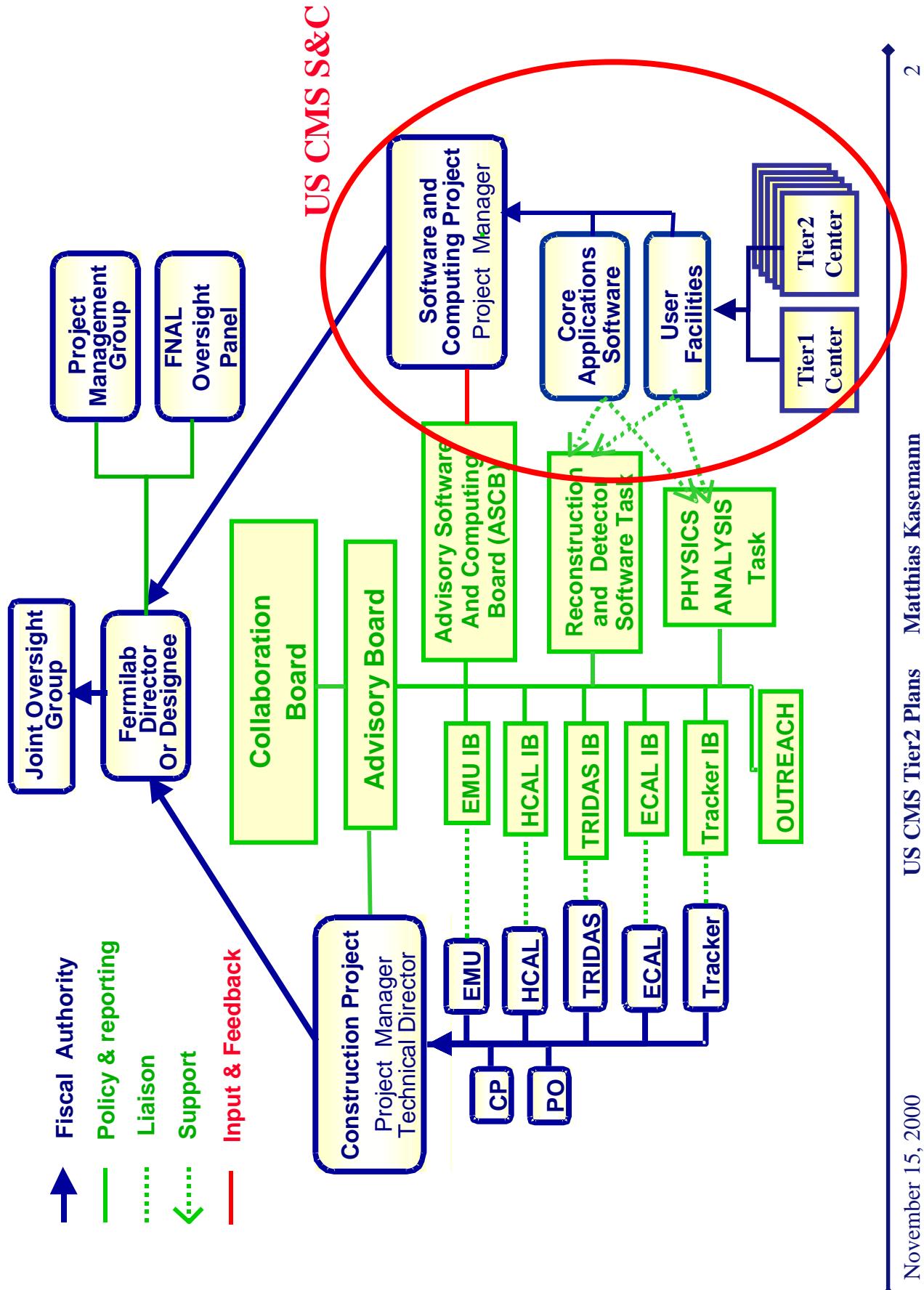
US CMS Software and Computing Tier 2 Center Plans

Matthias Kasemann
Fermilab

DOE/NSF Baseline Review
November 14-15, 2000
BNL



Organization Of US CMS Projects





Tier 2 Centers

- ♦ Tier 2s will partner with a Tier 1 which will provide support and liaison with CERN.

- ♦ US CMS Tier 2 Centers (or the CMS parts of multiple customer centers) appear as Level 3 WBS items under the UF project, and the UF Level 2 manager has specific responsibilities with respect to them
 - Tier 1 interfaces to the Tier 2's in the US
 - The Tier 2 CMS representative is a L3 manager of the US S&C project
 - Each Tier 2 signs MOU with US CMS C&C Project



Production Tier 2 Functions

- ◆ **Tier 2 (and Grid) philosophy**
 - ➔ Resource utilization decisions are closer to end users
 - ➔ Universities can leverage wider set of resources
- ◆ **Computing goals**
 - ➔ Simulation (inc. reconstruction)
 - ➔ User analysis
 - ➔ Component of CMS distributed analysis model
 - ➔ University support for analysis and simulation



Tier 2 Deliverables

- ◆ **The deliverables in both the prototype (and later in production) stages of the project are:**
 - demonstration and development of techniques and strategies for distributed computing that allow efficient exploitation of distributed resources including remote data management and data access tools.
 - demonstration that moderate size commodity computing clusters (of order 100 boxes) can be operated and managed with only minimal (<1.5 FTE) support.
 - computing cycles to contribute to CMS production (and analysis) activities
 - Strategy for work-sharing between Tiers
- ◆ **These categories of deliverables are inextricably intertwined since the distributed computing strategies cannot be verified without being applied to real world computing problems.**



Process to Decide Tier2S

- ◆ The location of Tier 2 centers will be decided by the project management in consultation with the US CMS collaboration and with CMS.
- ◆ We have already had some experience with this process in locating the first prototype Tier 2 center.
 - In this case, interest was solicited by the Acting L1 Project Manager and his decision was approved by the collaboration board.
- ◆ On the technical side, the management and collaboration should understand the role of each Tier 2 center.
 - It may be important to serve a well interconnected community of users.
 - It may be important to specialize on particular analysis topics.
 - An example of this could be heavy ion physics.
 - It may be important to leverage large computing resources that can be financed off project.
 - It may be important to minimize networking costs.
- ◆ It is too early to decide the exact criteria which will be telling in 2004-2006.



Tier 2 CPU Sizing: Simulation

- ♦ **Simulation: A “naturally” distributed activity**

- ♦ **Requirements**

- **Simulate 2×10^8 events per year for US analyses**
 - **1 event = 5000 SI95-sec (sim) + 3000 SI95-sec (recon)**
 - **Use 2×10^7 sec per year (65% efficiency)**

$$\text{Simulation (1 Tier 2)} = \frac{(2 \times 10^8) \times (8000 \text{ SI95-sec})}{5 \times (2 \times 10^7 \text{ sec})} = 16\text{K SI95}$$



Tier 2 CPU Sizing: Analysis

♦User analysis

- AOD and derived physics data \Rightarrow **20 Si95-sec**
- 200 physicists, 1% (10^7 events) data samples, 20 times/year
- Eff. $\approx 1/2$ simulation eff. \Rightarrow **less disciplined activity**
- Total Tier 2 analysis \approx **Tier 1 analysis**

$$\text{Analysis (1 Tier 2)} = \frac{200 \times (10^7 \text{ evts}) \times 20 \times (20\text{Si95-sec})}{5 \times (10^7 \text{ sec})} = 16\text{K Si95}$$

♦Total computing at Tier 2 = **16K + 16K = 32K Si95**

♦2006 CPU price estimate

- Doubling every 1.2 years $\Rightarrow \$4.7/\text{Si95} \times 32\text{K} = \150K
- Doubling every 1.5 years $\Rightarrow \$8.4/\text{Si95} \times 32\text{K} = \270K
- **\$150K - \$270K**



Tier 2 Disk Sizing

- ◆ **Simulation**
 - **Simulation needs satisfied by onboard (cheap) disk**
- ◆ **AOD/ESD**
 - Difficult to estimate, chaotic
 - Tier 2 is part of Grid, caches Tier 1 + Tier 0 data
 - 25 TB is rough estimate for average
- ◆ **Analysis**
 - Support several analyses (average 20% of US physicists)
 - Each analysis: $\approx 1\% = 10^7 \text{ events} \times 500 \text{ KB} = 5 \text{ TB}$
 - Total analysis $\approx 10 \times 5 \text{ TB} = 50 \text{ TB}$
- ◆ **Assume 50% efficiencies on these numbers**
- ◆ **Cost estimate**
 - $\$1.9K/\text{TB} \times 150 \text{ TB} = \$285K (\pm 25\% ?)$
 - $\$215K - \$355K$



Tier 2 Data Server Cost Estimate

- ♦ Data server
 - Runs database server ⇒ **substantial activity**
 - Need 20% of total CPU to serve data (**CERN + BaBar**)
 - Use actual CPU (no efficiency factors): **16K SI95**
- ♦ Data Server Cost
 - SMP prices
 - $0.20 \times 16K \times (\$18 - \$41) / SI95 = \$60K - \$130K$



Other Items

- ♦ **LAN switches**
 - ➔ 2000: 100 Mb/s switch for CPUs, GigE switch for uplinks
 - ➔ 2006: Same cost switches for higher performance
- ♦ **Small tape handling**
- ♦ **Software licenses**
- ♦ **Management tools**
 - ➔ Enterprise software for distributed analysis
 - ➔ Debugging, language
 - ➔ ...
- ♦ **Other items**
 - ➔ Infrastructure, installation, commissioning
 - ➔ Hardware maintenance (Data server, RAID system, ...)
 - ➔ Collaborative tools (Videoconferencing, remote vis., etc.)



Tier 2 1st Year Cost

- CPU Farm (32K \$195) \$ 150K - \$270K
- RAID Array (150 TB) \$ 215K - \$355K
- Data Server \$ 60K - \$140K
- LAN Switches \$ 60K
- Small Tape Library \$ 40K
- Tape Media and Consumables \$ 20K
- Installation & Infrastructure \$ 30K
- Collaborative Tools & Infrastructure \$ 20K
- Software licenses \$ 40K
- Maintenance \$ 0K *
- **Total Estimated Cost (First Year)** **\$635K – \$955K**

**Full Tier 2 cost (year 1): \$750K
Prototype Tier 2 cost (year 1): \$350K**

Assume
for now

* Maintenance is \$0K first year, \$30K per year afterwards



Tier 2 Yearly Costs

- Hardware evolution (3 year replacement) \$ 190K - \$300K
- Software licenses \$ 40K
- Maintenance \$ 30K
- Tape consumables \$ 20K
- **Total** \$ **280K - \$390K**

**Full Tier 2 yearly cost: \$250K
Prototype Tier 2 cost: \$120K**

- “Optimistic” for now
- Staff (Ops and System Support) \$ 200K
- Networking costs
- Average WAN connection (e.g., Abilene) \$ 300K(?)



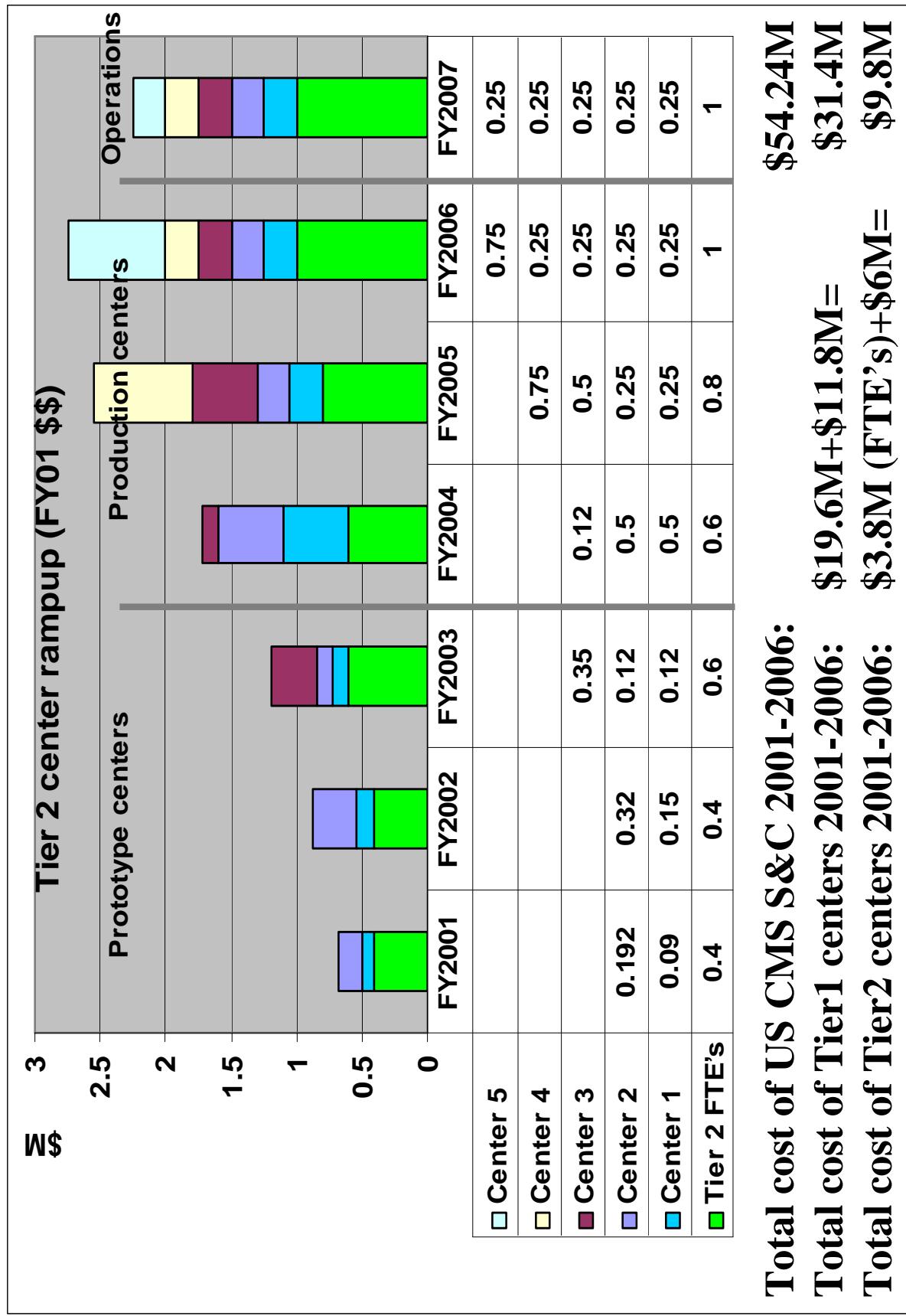
US CMS UF WBS: Tier 2 Centers

- ◆ Tier 2 centers are staffed to purchase, deploy, maintain and operate the hardware.
- ◆ Tier 2 center architecture and configuration will benefit from experience locally and at Tier0 and Tier1 sites.
- ◆ Tier 2 center are not expected to deal with mass storage systems.

	2001	2002	2003	2004	2005	2006	2007
1.4 Tier 2 Regional Centers	2.5	3.5	4	5	6.5	7.5	7.5
1.4.1 Tier 2 Center 1	1.5	1.5	1.5	1.5	1.5	1.5	1.5
1.4.1.1 T2 Design	0.25	0	0	0	0	0	0
1.4.1.2 T2 Equipment	0.5	0.5	0.5	0.625	0.625	0.625	0.625
1.4.1.2.1 Hardware Procurement & Installation	0.25	0	0	0.125	0.125	0.125	0.125
1.4.1.2.2 Commissioning	0.25	0.25	0.25	0.25	0.25	0.25	0.25
1.4.1.2.3 Hardware Maintenance	0	0.25	0.25	0.25	0.25	0.25	0.25
1.4.1.3 T2 Software Infrastructure Framework	0.5	0.375	0.5	0.5	0.5	0.5	0.5
1.4.1.3.1 Installation and Deployment	0.5	0.25	0.25	0.25	0.25	0.25	0.25
1.4.1.3.2 Maintenance	0	0.125	0.25	0.25	0.25	0.25	0.25
1.4.1.4 Integration with T1/T2	0.125	0.5	0.375	0.25	0.25	0.25	0.25
1.4.1.5 System Documentation	0.125	0.125	0.125	0.125	0.125	0.125	0.125
1.4.2 Tier 2 Center 2	1	1	1.5	1.5	1.5	1.5	1.5
1.4.3 Tier 2 Center 3	0	1	1	1	1.5	1.5	1.5
1.4.4 Tier 2 Center 4	0	0	0	1	1	1.5	1.5
1.4.5 Tier 2 Center 5	0	0	0	0	1	1.5	1.5



Tier 2 center plan





US CMS Tier2 planning

- ◆ **FY 2000/1** Build basic services; HLT milestones
1-2 prototype-Tier 2
Initial Grid system
work with Tier 1
- ◆ **FY 2002**
- ◆ **FY 2003/4** Tier 2 centers at second set of sites
 - ◆ **FY 2005/6** Tier 2 centers at last set of sites
 - ◆ **FY 2006** Production-quality Grid System
- ◆ **R&D systems:** leverage existing resources at Universities
◆ **Funding for Tier 2's to come mostly from NSF initiatives**



Tier2 Prototype Needs Now

- ◆ **Need to undertake R&D on distributed data replication and analysis now**
 - The time is right to attract computer science contributions to this problem
- ◆ **Serve Prototypical User Community to Serve Regional Analysis Needs**
 - Geographical Coherence: UCSD, Caltech, UCLA, UC Davis, UC Riverside
 - Work-Topic Coherence: Forward Muon Software; HLT Studies (with Florida)
 - Use this Sub-Community to help define and test modes of collaborative work on software and analysis
 - Use this Sub-Community to help define and test data analysis software needs
- ◆ **Take advantage of high speed networks in California**
 - Build analysis toolset and design/develop modes of collaborative work over networks
- ◆ **Production activities are occurring now, need non-centralized manpower**



First Tier2 Prototype Scale

◆ Requirements:

- Modestly sized for substantial distributed system R&D
 - 40-node cluster each in Caltech and UCSD; few TB disk arrays
- R&D activities demand participation in real live production efforts
 - (CMS will use 1,000 CPUs; 20 TB of data in Spring)
 - A Larger Production by Fall 2001 (PRS + S&C TDR)
- Balance of I/O-to-CPU for ORCA Production and Production-studies
- Caltech/UCSD Split over CALREN-2 and NTON (0.6 - 2.5 Gbps in 2001-2002)
- Leverage expertise, facilities, and high speed network equipment and infrastructure at San Diego, Caltech Center for Advanced Computing Research (and SLAC)
- Leverage Off-Project Software Engineers on PPDG, ALDAP, GriPhyN
- Given two major sites backing up the Tier2 development across a Gbps research network, equivalent tests could not be done at the Tier1 alone.



US CMS Tier 2 Center Decision

- ◆ Discussed at US CMS Collaboration Board, May 20, 2000
 - Agreed: start deployment of 1 prototype Tier 2 center
- Query sent to each US CMS institute:
 - Interest and plans to participate in Tier2 R&D now?
 - Who can work on it?
- Result:
 - Ready to started “now”: (Caltech + UC Davis + UCSD)9/00 Florida,
Univ. of Iowa + Iowa State ~4/01
- Other candidates for production:
 - Boston, Maryland, Minnesota Wisconsin
- Strategy: start first Tier2 center in California now
 - Substantial leveraging from local contributions (>25%)